DOCUMENT RESUME

ED 163 047 TM 008 075

AUTHOR Donlon, Thomas F.
TITLE Sex Differences in Test Speededness on the Scholastic Aptitude Test.
PUB DATE Oct 77
NOTE 27p.; Paper presented at the Annual Meeting of the Northeastern Educational Research Association (Ellenville, New York, October 26-28, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Aptitude Tests; *College Entrance Examinations; Conceptual Tempo; Females; Item Analysis; Males; Performance Factors; *Reaction Time; Response Style (Tests); Senior High Schools; *Sex Differences; Test Results
IDENTIFIERS *Scholastic Aptitude Test

ABSTRACT

Detailed item analysis results for a form of the Scholastic Aptitude Test were examined for evidence of sex differences in test speededness. The conclusions were: (1) there was no evidence of appreciable differences in rate-of-work on any section of the Scholastic Aptitude Test; (2) there was some evidence that low-scoring females on the mathematical sections work at faster rates than would be anticipated by the data for males or higher-scoring females; and (3) there was some evidence that while the time limits were so generous as to reduce variance in failure to complete, there may have been sizable differences in time to go back over the test; the score implications of this variance are seldom considered. The complexity of speed-power relationships is such that sex differences could possibly exist in a number of ways not considered in this study. Omitting an item, for example, as opposed to not reaching it, is often held to have rate-of-work implications. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED163047

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Sex Differences in Test Speededness on the Scholastic Aptitude Test

Thomas F. Donlon

Educational Testing Service

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Thomas Donlon

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND ORGANIZATIONS OF THE ERIC SYSTEM."

NERA
Ellenville, NY
October 26-28, 1977

TM008 075

Sex Differences in Test Speededness on the Scholastic Aptitude Test

Thomas F. Donlon
Educational Testing Service

Increasingly, there is an awareness among educators and in society at large of the restrictions which sex role stereotype can place upon the full development of human potential. As one element of this reexamination of traditional sex roles, the evidence concerning male-female differences in aptitudes must be carefully reexamined.

While studies concerning sex differences abound in the educational, psychological, and sociological literature, results have often been conflicting or inconclusive due to such methodological problems as overlapping distributions, confounding personality variables, and content bias within test items. Very different item types have been used in different tests which purportedly measure the same aptitude. Sex bias in test results may be introduced by yet another methodological factor--the degree to which test speededness differentially affects the performance of males and females on various aptitude tests. To the extent that speededness is operating differentially, it may modify the relationship between the sexes by either creating or masking differences.

A typical pattern of female superiority in verbal skills and male superiority in quantitative tasks has been consistently established. The reviews of Maccoby and Jacklin (1974) and Sherman (1971) are good examples. Much more research is, of course, required before we can assert whether these apparent differences are developmental in origin or the result of the methods by which we assess verbal and quantitative skills. To the extent that apparent sex differences are due to factors in the test and are not

reflective of general skill differences, a source of error in making future decisions about individuals has been introduced (Anastasi, 1958). Recent studies of sex differences in test performance have discussed item content factors as a source of sex bias (Hicks, Donlon and Wallmark, 1976); Strassberg-Rosenberg and Donlon, 1975). But an even more subtle bias may result from the differential effects of test speededness for the sexes. For example, if a male and a female have the same capacity for success on a test, but one works faster and hence answers more items, the faster-working student will appear to be more proficient in the ability tested. To the extent that speededness is operating differentially, it may modify the relationship between the sexes by either creating differences where none exist or masking existing differences.

Research evidence suggests that rate of response is more consistently associated with extraneous subject attributes, such as personality traits, than it is with the ability to respond correctly (Wesman, 1960). Among examinees who do not have time to attempt every item are those who have to consider carefully all or most of the choices in the items to weigh conscientiously their selections of correct responses. Many of these examinees possess sufficient information to logically determine a correct answer, but are unable to recognize the answer immediately. The additional time required by these students to process a response may in fact reflect a lack of practice with the task rather than a lack of ability (Moreton and Butcher, 1963; (Myers, 1952, 1958).

In the course of sex role development, males are reinforced for developing proficiency in masculine sex-typed aptitudes (e.g., mathematical

ability), while females are rewarded for practicing feminine sex-typed aptitudes (e.g. verbal ability). Hence if test speededness were operating differentially for the sexes, it would be expected that males would score lower on speeded tests of "feminine" aptitudes, while females would score lower on speeded tests of "masculine" aptitudes.

Kappy-Lewis and Donlon (1977) analyzed ten tests in the Cooperative Guidance and Placement Program comparing sex differences in rates of work for male and female samples of approximately 600 subjects each. There was little evidence of any difference. When the analysis was extended to a comparison of subgroups, defined by ability, some evidence of sex differences emerged. In two mathematical tests in particular it was found that the male subjects exhibited consistent speed and power hierarchies: the highest scoring ability group was the fastest working ability group. For females, however, the lowest scoring group was the fastest working.

The present study was undertaken to examine whether test speededness differentially affects the performance of males and females for the College Board Scholastic Aptitude Test (SAT), a widely known test used for selection for college entrance. This study extends the previous work by Kappy-Lewis and Donlon (1977). Differential patterns of speededness for varying ability levels across sex have also been found in a reanalysis of data on verbal and quantitative aptitudes supplied by McPeek (1970). In addition, Ford (1976) reported strong differences in test speededness between a selected sample of older women and a mixed-sex sample of community college students for reading and mathematical aptitude tests of the Comparative Guidance and Placement Program (CGP). All in all, it seemed prudent to examine the item analyses data for the SAT.

Method

Subjects. Spaced samples of 1,000 males and 1,000 females were drawn from the total group of SAT candidates who were tested on the SAT in April 1974. While the samples were intended to be representative of the general CEEB population, they were not restricted on any variable other than sex.

The item analysis program provided, in addition to the description of the total sample, an option for information on the performance of subgroups defined by total score. These groups were defined as the top 20%, the next to highest 20%, etc., five groups in all. For each such quintile the item analysis provides the basis for a computation of percent succeeding.

The groups were not equated prior to the analysis. This is not likely to produce any sampling considerations for the data for the SAT-V, for on this measure in recent years the sex differences have been negligible. This means that ability subgroups are quite comparable: the highest scoring fifth of males has, on the average, about the same score as the highest scoring fifth of females.

For SAT-M, however, the mathematical test, there are reported sex differences, so that each quintile for males is likely to be somewhat superior to the corresponding quintile for females. To the extent that rate of work is correlated with ability, one would expect that the males at a given quintile level would show slightly less drop-out as they worked through the test, when compared to their female counterparts. The sex difference, however, while appreciable is by no means so great as to make

the comparisons by quintile meaningful. In the absence of a cost-efficient technique for equating groups, the samples were accepted as drawn.

Instruments. The SAT at the time of this study consisted of the two basic tests, SAT-V and SAT-M, each further subdivided into two sections. Table 1 shows the number of items, time limit, and items/minute necessary to complete each section of the test.

The content of the two verbal sections was of four item types: analogies, sentence completion, antonyms and passage-related reading comprehension questions. The content of the two mathematical sections was general mathematical material and was similar except that Section IV included data sufficiency. This is a simpler item type, calling for a response indicating not what the answer is but whether there is enough information to calculate it. On this basis, Section IV would be anticipated to be somewhat less speeded. It has been established in earlier studies that this item type is handled more rapidly by the candidates.

- - - - - - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - - - - - -

Results

The results are primarily presentable in the form of a series of figures. This reliance upon graphic presentation, without the use of associated tables is considered justified by the essentially straightforward nature of the data. Thus, the gist of the results are developed in Figures 1 - 4 and 5 - 9, which show comparison of dropout rates for males and females overall and within ability levels.

Figure 1 - 4 describe total group comparisons. Each figure is relevant to a given section of the SAT. Thus, Figure 1 describes the

the dropout configurations for the first verbal section. It is evident that the experience of the groups is very similar throughout the entire range of the plot: this pattern is essentially retained in Figures 2 - 4. Overall, for the total groups, there is little evidence of any difference, and certainly no evidence of any difference which migh. lead to a score bias. There is a noticeable gap between the male and female dropout curves on the Mathematics 4 section, but in absolute terms this difference is inconsequential.

The tests are very little speeded, but the shorter timed sections, the 30-minute Verbal 1 and the 30-minute Mathematical 3, are more speeded than their 45-minute counterparts.

- - - - - - - - - - - - - - - - - - - -

Insert Figures 1 - 4 about here

- - - - - - - - - - - - - - - - - - - -

The plots are developed on normal probability paper. This practice derives from some recent work by Donlon and Reilly (1976) on the hypothesis that rate-of-work on multiple choice tests is a normally distributed variable. The graphic technique permits a study of a number of characteristics that are difficult to capture through statistical tests of tabular data. For example, the basis hypothesis of normality can be assessed by considering the linearity of the lines. While all plots show some deviations, fairly good-fitting straight lines could be drawn through the plots in Figures 1, 2 and 3. Thus, there is some evidence of a basic normal distribution in these figures.

Additional interpretations offered by the plots are the hypothetical mean number of items which can be reached in the time allowed, and the

standard deviation of the number of items which can be reached. It is evident that many more items can be reached than are required. Table 2 offers the hypothetical values for each section

- - - - - - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - - - - - -

The values in Table 2 suggest that some individuals in Mathematics would be able to do over half again as many items as they are being required to. While there are no sex differences in these data, they do reflect a related problem: do the sexes use their "time left over" in the same ways?

Still another feature of such graphs, though not linked to sex differences, is shown by the discontinuity in the graph for Verbal 1. To indicate this discontinuity clearly, Figure 1 has been redrawn as Figure 5, with straight lines added by a visual fit. Clearly the move from Item 35 to Item 36 is a discontinuity, and there is evidence of a similar shift at the end, for Item 40. In each case, the number attempting Items 36 and 40 is less than one would expect, given the previous trends. The explanations, however, are somewhat different.

- - - - - - - - - - - - - - - - - - - -

Insert Figure 5 about here

- - - - - - - - - - - - - - - - - - - -

Item 36 is preceded by a passage. When time was called, some persons were still reading the passage. They did not attempt the first item in the related set, Item 36. Item 40, however, is the last item in the test;

apparently people who were working on it when interrupted did not make a response.

These findings of discontinuities do not counter the fundamental hypothesis of normality. The linear fits in Figure 5, in spite of the discontinuity at Item 36, are similar, suggesting a basic work rate, in the sense of items/minute, which is necessarily deviated from when a passage is encountered but which is resumed after it is read. It also tends to indicate that people do read the passage before attempting items.

Figures 6 - 9 extend the quest for viable sex differences to the comparisons within ability groups. In the interest of visually manageable graphs, only two of the five ability groups defined by the item analysis are presented. These are the highest ability males and females, the top 20% of each of their respective distributions, and the lowest 20%. It is evident from these figures that work rate is related to overall success. The ability groups are defined on total score, and they show, not surprisingly, that the highest scoring groups attempt more items, on the average.

The patterns of difference between males and females are not pronounced in this analysis. The lines do not coincide, but they show an essential similarity that indicates that there is no difference great enough to be a significant factor in the observed raw score differences. In point of fact, the observed raw score differences for the total test are more marked for mathematics than they are for the verbal sections. The raw score means for the sexes on verbal are virtually equivalent. But for mathematics, this is not the case, with males doing somewhat better.

- - - - - - - - - - - - - - - - - - - -

Insert Figures 6 - 9 about here

- - - - - - - - - - - - - - - - - - - -

This situation makes the patterns in Figure 8, and to a certain extent in Figure 9, somewhat remarkable. In each case the low-scoring females are finishing more of the test than the low scoring males. But the effect is so pronounced for Mathematics 3 that the low scoring females show less dropout than the high scoring females. In other words, more of the poorest scoring females are answering the last question than the highest scoring females. The difference is slight, but when one considers where these low-scoring females "ought" to be, it is apparent that they are working at a much faster rate than would be anticipated.

This finding is consistent with earlier informal data supplied by Wallmark and Conrad (1975) where Black females were low-scoring but high finishing, and with the results of Kappy-Lewis and Donlon (1977) where there were no clear-cut speed distinctions among ability levels for females on mathematics.

Another approach to these phenomena is available through considering not the percentage still working, which is the basic measure in Figures 1 - 9, but the mean ability of those still working. This information is available for the total groups of males and the total groups of females through the standard ETS item analysis of the SAT. Presumably the mean ability of those remaining within quintiles is also affected, but no information on these special subgroups was available.

The data for the total group is presented in Table 3. In each case the mean is arbitrarily established as 13.0 for the total group.

The means of any items earlier than those described in the Tables were 13.0 for both males and females. Thus, each subtable begins with the last item for which both sexes had means of 13.0 and proceeds to the last item.

- - - - - - - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - - - - - - -

The typical pattern is a rise in mean ability as dropout occurs. Thus, speed and score seem to be correlated in a logical sense: those who drop out get lower scores. But for the females on Mathematics 3 there is a decline in average ability, indicating that those who drop out are nonetheless receiving higher scores than those who remain. This phenomenon is in a sense another reflection of Figure 8. It suggests that the work rates of low scoring females on mathematics may be quite unique. No firm rationale for this finding can be advanced. It may be influenced by the composition of this section, which included data sufficiency items. It is in a sense an indication that while the sexes are highly comparable in any total group sense there may be a female subgroup which differs somewhat from males and from their female counterparts.

Summary

The complexity of speed-power relationships is such that sex differences could possibly exist in a number of ways not considered in this study. Omitting an item, for example, as opposed to Not Reaching it,

is often held to have rate-of-work implications. But in general, the following conclusions appear justified:

1. There is no evidence of appreciable differences in rate-of-work on any section of the Scholastic Aptitude Test.
2. There is some evidence that low-scoring females on the mathematical sections work at faster rates than would be anticipated by the data for males or higher-scoring females.
3. There is some evidence that while the time limits are so generous as to reduce variance in failure to complete, there may be sizable differences in time to review, go back over the test, etc., the score implications of this variance are seldom considered.

References

Anastasi, A. Differential Psychology (3rd ed.). New York: The Macmillan Company, 1958.

Donlon, T. F., & Reilly, R. R. Estimating speededness parameters on several tests. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1976.

Ford, S. F. Test speededness for subjects without recent school experience. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1976.

Hicks, M. M., Donlon, T. F., & Wallmark, M. Sex differences in item responses on the Graduate Record Examination. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, 1976.

Kappy-Lewis, K. & Donlon, T. F. Test speededness as a factor in sex differences. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, April, 1977.

Maccoby, E. E., & Jacklin, C. N. The psychology of sex differences. Stanford, CA: Stanford University Press, 1974.

McPeek, W. M. Experimental Pretests in RSA45. Internal ETS Memorandum, 1970.

Moreton, C. A., & Butcher, H. J. Are rural children handicapped by the use of speeded tests in selection procedures? British Journal of Educational Psychology, 1963, 33, 22-30.

Myers, C. T. The factorial composition and validity of differently speeded tests. Psychometrika, 1952, 17, 347-353.

Myers, C. T. Some observations of problem solving in spatial tests. Research Bulletin 58-16. Princeton, NJ: Educational Testing Service, 1958.

Sherman, J. A. On the psychology of women: A survey of empirical studies. Springfield, IL: Charles C. Thomas, 1971.

Strassberg-Rosenberg, B., & Donlon, T. F. Content influences on sex differences in performance on aptitude tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Washington, D.C., 1975.

Wallmark, M. M., & Conrad, L. Report on the item analysis of a GRE aptitude test by ethnic and sex subgroups. Internal ETS Memorandum, 1975.

Wesman, A. G. Some effects of speed in test use. Educational and Psychological Measurement, 1960, 20, 267-274.

Table 1

Rate of Work Requirements for Each Section of the SAT

Section	Number of Items	Time Limit	Items/Minute Necessary to Complete	Seconds/Item
I (Verbal)	40	30	1.33	45
II (Verbal)	50	45	1.11	54
III (Mathematical)	25	30	0.83	72
IV (Mathematical)	35	45	0.77	77

Table 2

Estimated Mean and Standard Deviation
of the Number of Items Which can be Reached

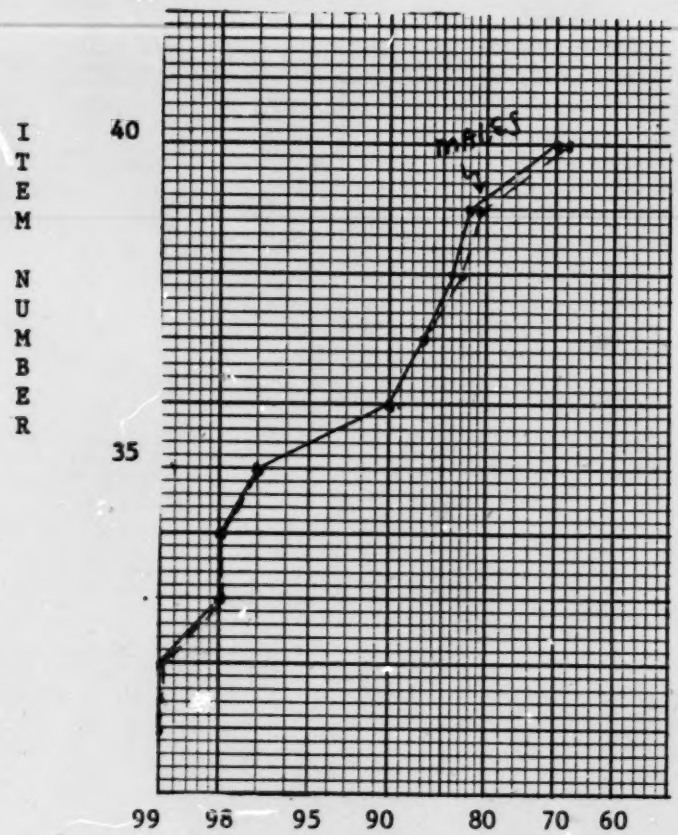
Section	Number of Items	Estimated Mean	Estimated Standard Deviation
Verbal 1	40	42	5
Verbal 2	50	58	8
Mathematics 3	25	29	5
Mathematics 4	35	46	8

Table 3

Average Ability for Males and Females
for Final Items in Four Subtests of the SAT

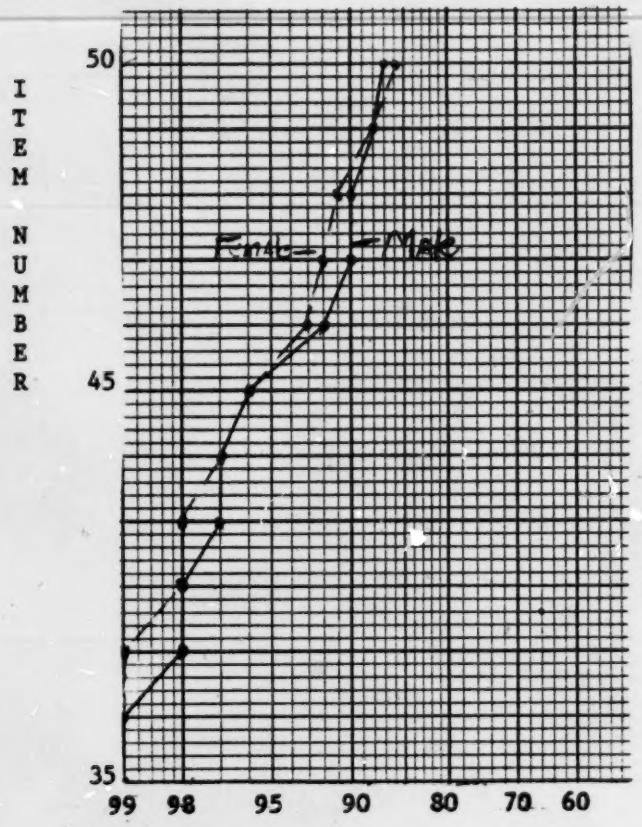
Verbal 1			Verbal 2			Mathematics 3			Mathematics 4		
Item	Male	Female	Item	Male	Female	Item	Male	Female	Item	Male	Female
32	13.0	13.0	40	13.0	13.0	20	13.0	13.0	29	13.0	13.0
33	13.1	13.0	41	13.0	13.1	21	13.1	13.1	30	13.1	13.0
34	13.1	13.0	42	13.1	13.1	22	13.1	13.1	31	13.1	13.1
35	13.1	13.1	43	13.1	13.1	23	13.1	13.0	32	13.1	13.1
36	13.3	13.2	44	13.1	13.1	24	13.1	13.0	33	13.2	13.1
37	13.3	13.3	45	13.1	13.1	25	13.1	12.9	34	13.2	13.1
38	13.3	13.3	46	13.2	13.2				35	13.2	13.2
39	13.4	13.3	47	13.2	13.2						
40	13.5	13.4	48	13.2	13.2						
			49	13.3	13.3						
			50	13.3	13.3						

Figure 1



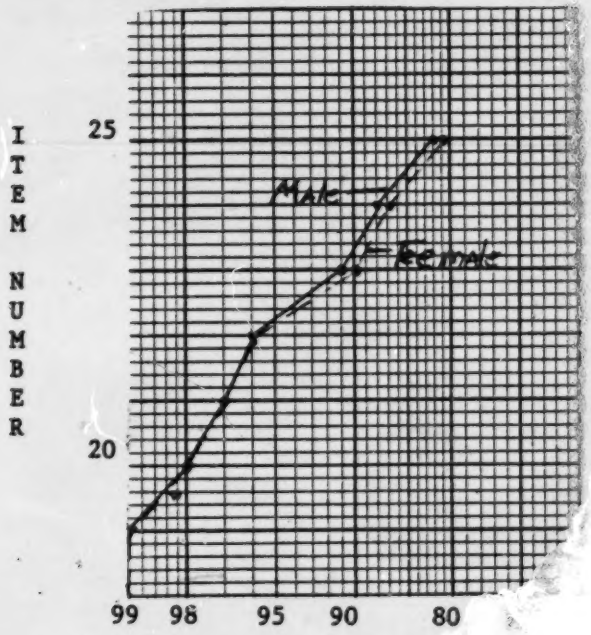
% Completing Item
Verbal Section 1
Males and Females
Total Group

Figure 2



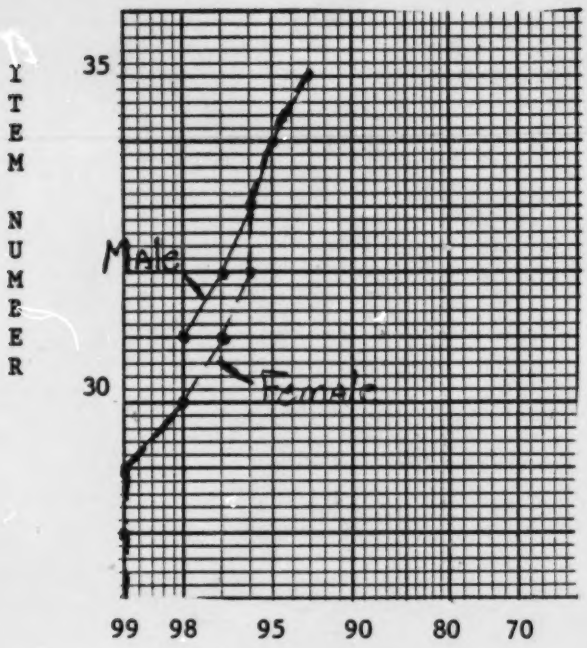
% Completing Item
Verbal Section 2
Males and Females
Total Group

Figure 3



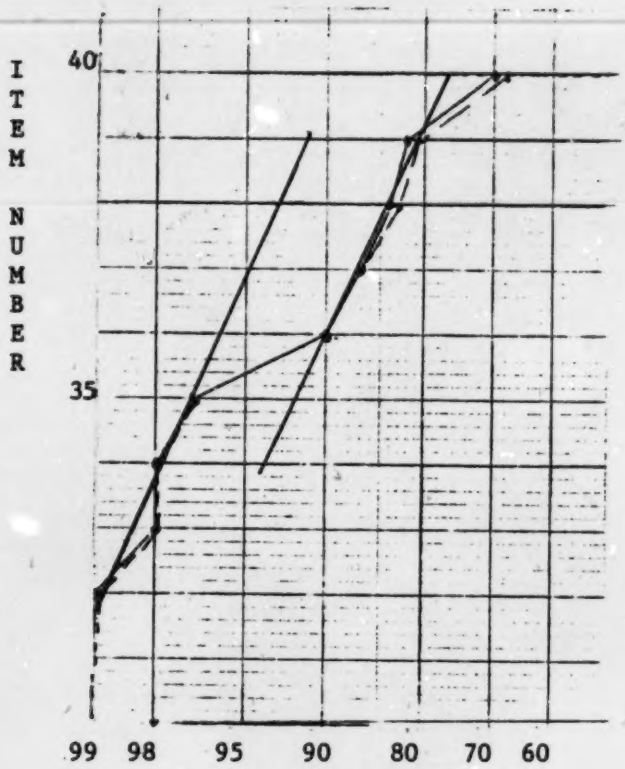
% Completing Ite
Mathematical Sect
Males and Females
Total Group

Figure 4



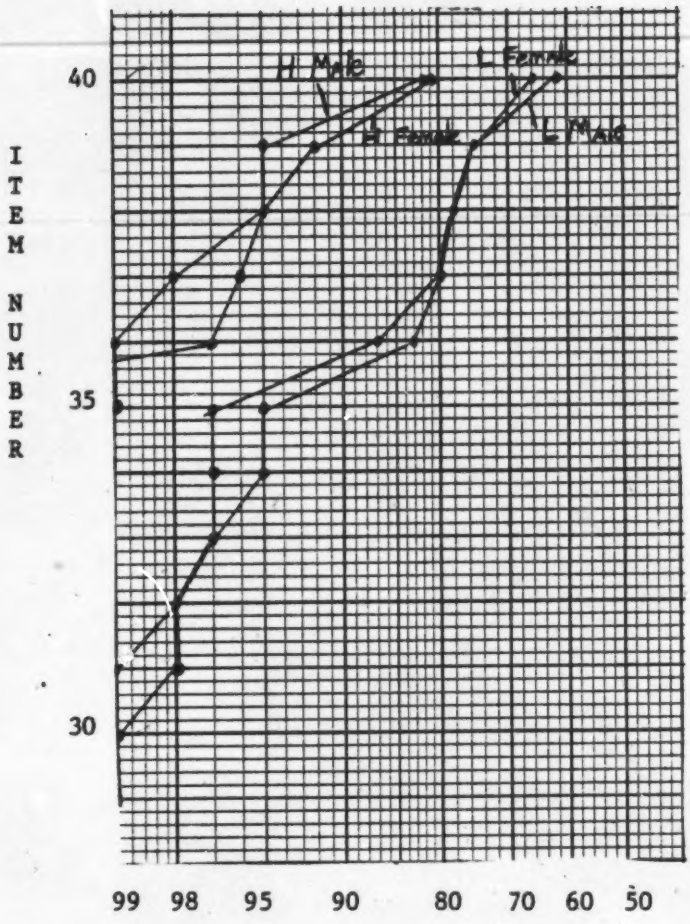
% Completing Item
Mathematical Section 4
Males and Females
Total Group

Figure 5



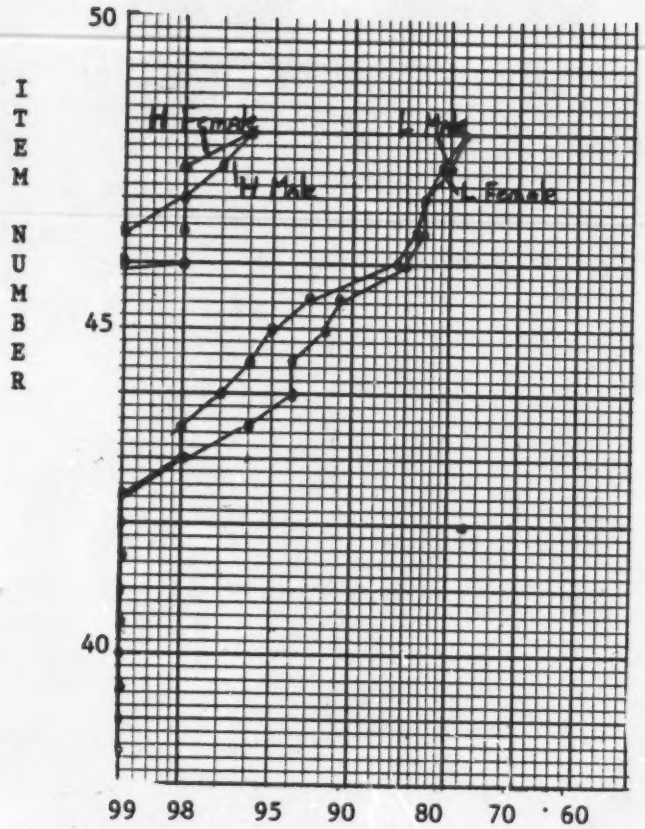
% Completing Item
Verbal Section 1
Males and Females
Total Group

Figure 6



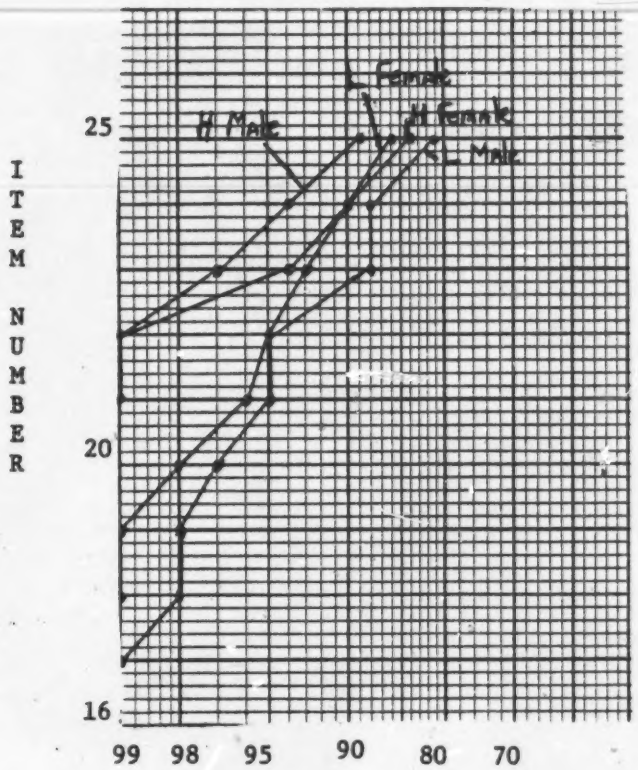
% Completing Item
Verbal Section 1
Males and Females
Highest and Lowest
Quintiles

Figure 7



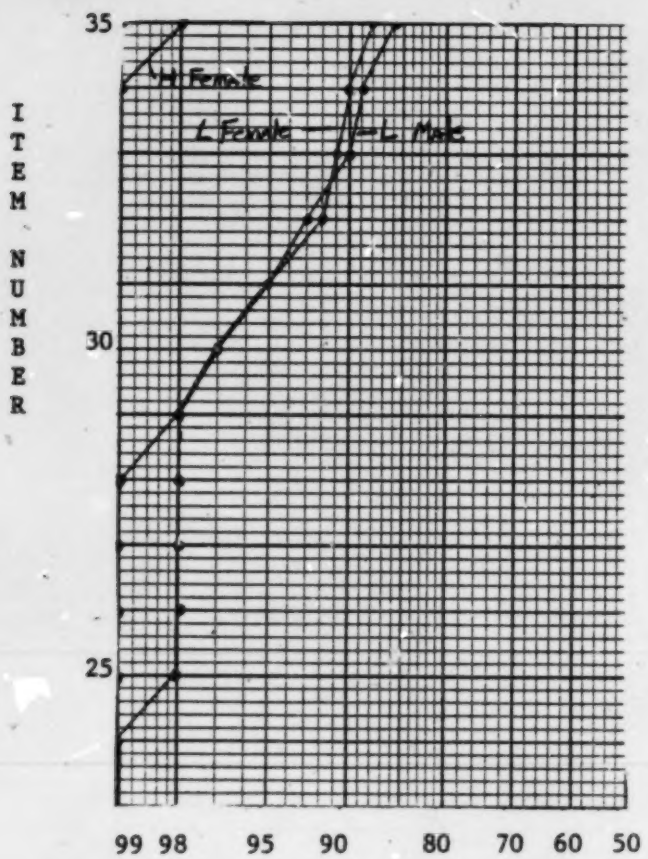
% Completing Item
Verbal Section 2
Males and Females
Highest and Lowest
Quintiles

Figure 8



% Completing Item
Mathematical Section 3
Males and Females
Highest and Lowest
Quintiles

Figure 9



% Completing Item
Mathematical Section 4
Males and Females
Highest and Lowest
Quintiles